


Gerald Amiel Ballena

✉ gmballena@up.edu.ph  github.com/GABallena  linkedin.com/in/gerald-amiel-ballena

📍 Metro Manila, Philippines

 orcid.org/0009-0000-8857-9755

Summary

Skills

- **Programming:** Python, Bash, R, Snakemake, Conda, LaTeX
- **Bioinformatics Tools:**
- **Data Visualization:** ggplot2, Krona, Pavian, Matplotlib, Plotly, Shiny
- **Databases:** UniProt, CARD, NCBI, KEGG, RefSeq, SILVA, BUSCO
- **Operating Systems:** Ubuntu, Windows
- **Version Control:** Git, GitHub

Experience

Project Technical Specialist

University of the Philippines, College of Public Health

2024 – Present

- Developed dynamic metagenomics pipelines for AMR detection.
- Spearheaded the creation of custom data visualization scripts using R and Python.

Education

M.S. Microbiology

University of the Philippines Diliman

Graduated: 2022

Thesis:

Specifics:

B.S Biology

University of the Philippines Baguio

Graduated: 2018

Thesis:

Specifics:

Projects (See GitHub)

Project4

GitHub: Project4

- most is NDA
- Identification of ARG-associated contigs via iterative alignment of reverse-translated CARD k-mers

Automated Read Trimming and Quality Control Pipeline

GitHub: Project4

- Engineered a dynamic pipeline integrating **Trimmomatic**, **Fastp**, **BBDuk**, and **Cutadapt** with randomized trimming parameters across iterations, optimizing read preprocessing through extensive parameter testing.
- Automated **FastQC** assessments with retry logic for failed samples, ensuring robust quality control and traceability via **TSV logs** and **YAML-driven configurations**.
- Streamlined the workflow with comprehensive summary reporting on sample quality and trimming performance, enabling scalable, reproducible analyses in high-throughput genomic projects.

Side

GitHub: Side

- Pioneered an advanced **k-mer sequence space exploration pipeline**, enabling the detection of **biological signals** while effectively filtering out metagenomic noise to enhance data accuracy.

- Engineered **optimization of trimming parameters** using a **Metropolis-Hastings algorithm (MCMC)** to systematically refine high-dimensional genomic data for superior precision in downstream analyses.
- Implemented a **wavelet-based normalization strategy** on k-mer counts using **Continuous Wavelet Transform (CWT)**, ensuring robust signal processing and filtering of repetitive regions to spotlight high-coverage k-mers.
- Developed a high-throughput **automated k-mer extraction system** leveraging **nested dictionaries** and **pandas** to dynamically structure and organize k-mer counts by SCG, streamlining large-scale data management.
- Generated comprehensive **variance analysis reports**, driving deeper insights into the genomic landscape by identifying **conserved regions** and **highly variable loci**, propelling metagenomic research forward.

Main

GitHub: Main

- Developed an **automated version control system** for managing bioinformatics tools, dynamically tracking and updating software packages from the **Bioconda** repository to ensure up-to-date environments across projects.
- Leveraged the **Anaconda API** to automate the comparison of installed tool versions with the latest Bioconda releases, seamlessly updating Conda YAML files and improving reproducibility in bioinformatics pipelines.
- Built a framework for parsing **Conda and pip dependencies**, simplifying YAML file management by automating updates and reducing manual intervention in large-scale computational biology projects.
- Automated version control of dependencies within wrapper

Publications

- Ballena., et al.

Certifications

- Coursera

Certifications

- Nextflow for Bioinformatics, Coursera (2022)
- Machine Learning for Genomics, edX (2021)
- Data Visualization with Python (Plotly, D3.js), Udemy (2021)
- HPC for Bioinformatics, Coursera (2020)

I hereby certify that the information provided above is true and correct to the best of my knowledge.

Gerald Amiel M. Ballena
Signature